Received: 9 August 2018 | Revised: 11 December 2018 | Accepted: 13 December 2018

DOI: 10.1002/yea.3374

SPECIAL ISSUE ARTICLE

WILEY Yeast

The problem of genetic code misreading during protein synthesis

Kartikeya Joshi* | Ling Cao† | Philip J. Farabaugh

Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, Maryland, USA

Correspondence
Philip J. Farabaugh, Department of Biological Sciences, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21228, USA.
Email: farabaug@umbc.edu

Funding information
National Institutes of Health, Grant/Award Number: GM 029480; National Science Foundation, Grant/Award Number: 1645795

Abstract

Saccharomyces cerevisiae has been an important model for determining the frequency of translational misreading events, those in which a tRNA pairs incorrectly to the mRNA and inserts an amino acid not specified by the codon in the mRNA. Misreading errors have been quantified in vivo using reporter protein systems or mass spectrometry with both approaches converging on a simple model for most misreading. The available data show that misreading tRNAs must form stereotypical base mismatches that correspond to those that can mimic Watson–Crick base pairs when formed in the ribosomal A site. Errors involving other mismatches occur significantly less frequently. This work debunks the idea of an average misreading frequency of 5×10^{-4} per codon that extends across the genetic code. Instead, errors come in two distinct classes—high frequency and low frequency events—with most errors being of the low frequency type. A comparison of misreading errors in *S. cerevisiae* and *Escherichia coli* suggests the existence of a mechanism that reduces misreading frequency in yeast; this mechanism may operate in eukaryotes generally.

KEYWORDS

Saccharomyces, translational error, misreading, tRNA modification

1 | INTRODUCTION

From the earliest speculation on the nature of the genetic code, it has been axiomatic that each codon must correspond to a single amino acid (Crick, Barnett, Brenner, & Watts-Tobin, 1961). But misreading events are common enough that a substantial minority of proteins produced can include at least one error (Parker, 1989). On the basis of his work with enzymes discriminating similar substrates, Pauling (1957) suggested that proteins could be "statistical," meaning that they might carry many random substitutions of amino acids not encoded in the messenger (Woese, 1965). Pauling estimated a maximum error rate as high as 5%, but many experiments have shown that amino acid misincorporation errors are at least 100-fold less frequent (reviewed in Bullwinkle, Lazazzera, & Ibba, 2014). Protein synthesis employs multiple strategies to amplify accuracy. Kinetic proofreading (Hopfield, 1974; Ninio, 1975) was proposed as a means to reuse discrimination to achieve higher accuracy. tRNAs bind to the ribosome's decoding center as a ternary complex of an elongation factor (EF-Tu in bacteria or EF-1A in eukaryotes), GTP, and aa-tRNA. Complexes with near-cognate tRNAs (those making one mismatch with the codon) are discriminated against during initial selection then again after an irreversible GTP hydrolysis step (Thompson, 1988). More recently, ternary complex discrimination has been broken down into many well-defined kinetic steps (Rodnina, Fischer, Maracci, & Stark, 2017). The most important for accuracy are activation of the EF-Tu/EF-1A GTPase and accommodation of the aminoacyl end of the tRNA into the ribosome's peptidyl transferase center both of which are significantly faster for cognate rather than near-cognate ternary complexes (Rodnina & Wintermeyer, 2001; Wohlgemuth, Pohl, & Rodnina, 2010). The details of the discrimination process, however, remain controversial (Pavlov, Liljas, & Ehrenberg, 2017). Discrimination involves an induced fit of the cognate codon–anticodon complex in the ribosomal A site that stimulates a large-scale rearrangement of

*Personal Genome Diagnostics, Baltimore, MD 21224.
†US Food and Drug Administration, Silver Spring, MD 20993

the ribosome, which effectively traps the cognate ternary complex on the ribosome (Fischer et al., 2016; Loveland, Demo, Grigorieff, & Korostelev, 2017; Ogle, Murphy, Tarry, & Ramakrishnan, 2002). The details of how some near-cognate tRNAs can effectively compete for acceptance by undermining this selective scheme are now the area of greatest interest in this field and will be a major subject of this review.

Data on the kinetics of aa-tRNA selection come from *Escherichia coli*, but the process of translation elongation in prokaryotes and eukaryotes is extremely similar. There is every expectation that the kinetic mechanisms in *Saccharomyces cerevisiae* are also extremely similar. The frequency of misreading errors has been mainly studied in *E. coli*, but recent work in yeast has closed the gap in understanding and demonstrated that again prokaryotes and eukaryotes are very similar (Blanchet et al., 2018; Blanchet, Cornu, Argentini, & Namy, 2014; Joshi, Trivedi, & Farabaugh, 2018; Kramer, Vallabhaneni, Mayer, & Farabaugh, 2010). Neither of these observations are surprising; the most important structures and processes of translation appear to have been established at the time of the last universal common ancestor. Some processes have diverged with the eukaryotic version becoming more elaborate (translation initiation, peptide termination), but elongation has not changed significantly. We will therefore include discussion of bacterial data where appropriate in this review.

2 | A SUBSET OF MISREADING ERRORS ARE HIGHLY EFFICIENT

Early studies of misreading frequencies in *E. coli* produced data showing a variety of misreading frequencies, as high as 5×10^{-2} per codon to as low as 5×10^{-5} per codon (reviewed in Parker, 1989). The variation in methodologies and the lack of a comprehensive approach made it difficult to evaluate what this variation indicated about error mechanisms. Despite the variation, Parker (1989) proposed an "average" error frequency of 5×10^{-4} per codon, which was seized upon by the field. Using a Poisson distribution assuming this level of error per codon, he suggested that for an average protein of 300 amino acids, about 14% of the proteins produced would include at least one error. In deriving an overall average, he assumed that the errors that had been quantified were representative of errors across the code. But, was this conclusion warranted?

This issue remained unexplored for about two decades, but in recent years, interest has been rekindled. This resurgence derives, we suspect, from increased concern for accurately proteins for biotechnology purposes. Protein overproduction can induce a disturbingly high frequency of errors in bacteria (Brinkmann, Mattes, & Buckel, 1989; Calderone, Stevens, & Oas, 1996; Seetharam et al., 1988). Similar overexpression experiments in CHO cells produced far less frequent errors (Wen et al., 2009; Yu et al., 2009), probably because the proteins were not as highly overproduced. The ubiquity of mistakes during protein expression for biotechnological purposes identify misreading as a problem that needs to be addressed if the quality of these proteins is to be optimized.

Mitigating such errors requires a better understanding of their causes. Available data on misreading reveal that errors are skewed toward events involving a few specific base mismatches (Blanchet et al., 2014; Joshi et al., 2018; Manickam, Nag, Abbasi, Patel, & Farabaugh, 2014; Z. Zhang, Shah, & Bondarenko, 2013). The most efficient misreading events involve one of three base mismatches: G•U, U•U, and C•U (bases shown in codon•anticodon order; Figure 1). Frequent errors have been observed for all tRNAs with a middle position U_{35} in the anticodon mismatching with a middle position G_2 in the codon (G_2•U_{35} mismatches) in *E. coli* (Manickam et al., 2014) and in *S. cerevisiae* (Joshi et al., 2018). G_2•U_{35} mismatch errors were independently observed by mass spectrometry in *E. coli* and mouse CHO cells (Z. Zhang et al., 2013). Errors have also been identified involving G_1•U_{35} mismatches by tRNAs that read cognate codons beginning with A misreading codons beginning in G. In fact, nine of the 11 possible G•U errors have been observed suggesting that the ability to form this mismatch strongly predisposes a tRNA to misread.

Many cases have been identified of errors involving pyrimidine–pyrimidine mismatches. Wobble U_3•U_{34} was identified in all cases tested, and wobble C_3•U_{34} mismatches in several (Kramer & Farabaugh, 2007; Manickam et al., 2014; Z. Zhang et al., 2013). Two U_1•U_{36} mismatches were also observed, but none involving C_1•U_{36} mismatches (Kramer & Farabaugh, 2007; Manickam et al., 2014). A single U_2•U_{35} mismatch was observed only under error-inducing conditions (Kramer & Farabaugh, 2007), but no C_2•U_{35} mismatches.

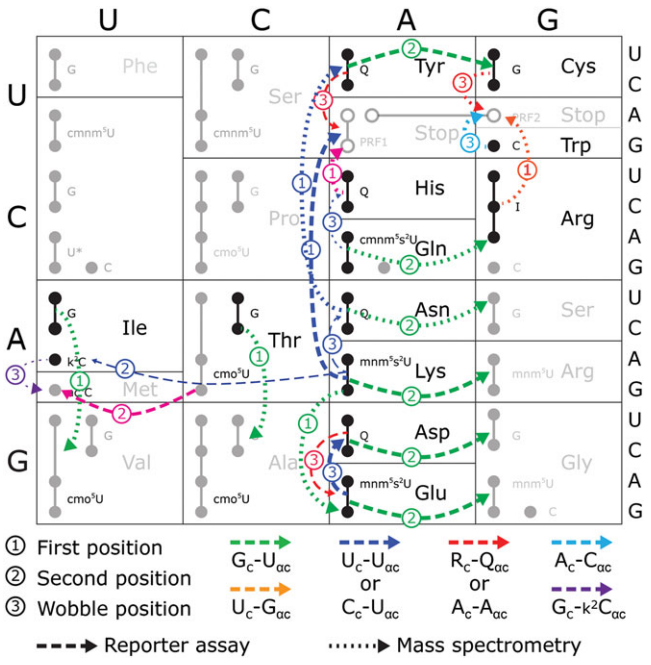

FIGURE 1 Only a subset of all possible misreading errors have been identified in vivo. The pattern of misreading by individual tRNAs is shown with reference to the set of tRNAs expressed in *Escherichia coli*, where much of this work has been done. Each codon is represented by a small circle; filled circles connected by lines are all recognized by a single tRNA, with the anticodon wobble nucleotide of each tRNA identified immediately to the right using the standard convention (Limbach, Crain, & McCloskey, 1994). Arrows connect the misreading tRNA and the codon(s) misread. The thickness of the arrow indicates relative frequency of the events (thick >10–4 per codon; thin <10–4 per codon). The arrows are labeled to indicate the position of the mismatch and color coded for the nature of the mismatched base pair, as shown [Colour figure can be viewed at wileyonlinelibrary.com]

These data show that a C•U mismatch is only tolerated at the wobble position and a U•U mismatch mostly at the first and wobble positions.

Importantly, almost all other mismatches induce errors with only very low efficiency, no more than about 2×10^{-6} per codon, whereas errors involving the three mismatches described above are much more frequent, up to 3.6×10^{-3} per codon in bacteria (Kramer & Farabaugh, 2007; Manickam et al., 2014) and 7×10^{-4} per codon in yeast (Joshi et al., 2018; Kramer et al., 2010). The 100 to 1,000-fold difference in apparent error frequency may underestimate the range of errors because the low activity recorded for most codons results from apparent enzyme activities near the background of the assay. In vitro analysis of misreading errors by $tRNA^{Glu}_{UUC}$ suggest the frequency on some codons may be as low as 2×10^{-8} per codon (J. Zhang, Ieong, Mellenius, & Ehrenberg, 2016). These data suggest that there is a difference in kind between the codon–anticodon interaction of a tRNA susceptible to high levels of error and those that are highly accurate. It is arguably more generally relevant that the frequency of errors at most codons is so low. The mechanism of error correction apparently is remarkably effective—far more effective than could be justified by any imaginable selective pressure. It would appear, therefore, that the need to suppress certain errors was great enough to enforce a system that restricted these errors as much as possible, and as a result, the system nearly quantitatively eliminates all other errors. The categorical nature of most error elimination is arguably unexpected and should fundamentally alter our perception of error correction during translation.

3 | THE MOST FREQUENT ERRORS ARE CAUSED BY MOLECULAR MIMICRY

The last few years have seen a large increase in our understanding of the process of aa-tRNA recruitment and of the origins of misreading errors. Some of the original work on these topics proposed that cognate but not near-cognate ternary complex can efficiently bind the A site and induce a large-scale rearrangement of the ribosome called domain closure that both prevents dissociation of the ternary complex and activates the EF-Tu GTPase (reviewed in Rodnina et al., 2017). This large-scale rearrangement involves induced fit in which conformational changes in constituents of the A site, principally rRNA nucleotides G530, A1492, and A1493, allow them to contact all three base pairs of the codon–anticodon complex (Ogle et al., 2002). The near-cognate complexes tested were not able to induce these changes, suggesting that this inability explained the preference for cognate complexes. The details of this process have been further clarified first by X-ray crystallography of several cognate and near-cognate ternary complexes in the bacterial 70S ribosomal (Demeshkina, Jenner, Westhof, Yusupov, & Yusupova, 2012; Demeshkina, Jenner, Westhof, Yusupov, & Yusupova, 2013). This work confirmed the roles of the three rRNA nucleotides but showed that some near-cognates both induce domain closure and formed interactions with the rRNA nucleotides indistinguishable from those with cognates. These interactions require the mismatched nucleotides to adopt Watson–Crick (WC) geometry. A recent study using ensemble cryo-electron microcopy identified three distinct steps in the process of forming this A site complex again by both cognate and near-cognate ternary complexes (Loveland et al., 2017). The three steps involve the ternary complex increasingly entering deeper into the A site and the three rRNA nucleotides increasingly adopting the final structure interacting with the codon–anticodon complex.

Only G•U (Demeshkina et al., 2012; Rozov, Westhof, Yusupov, & Yusupova, 2016), U•G (Rozov et al., 2018), and U•U mismatched nucleotides (Rozov, Demeshkina, Westhof, Yusupov, & Yusupova, 2015) were found to adopt WC geometry; C•U pairs have not been yet analyzed. A•A and A•C pairs formed non-WC pairs of similar geometry (Rozov et al., 2015). As illustrated in Figure 2, although the

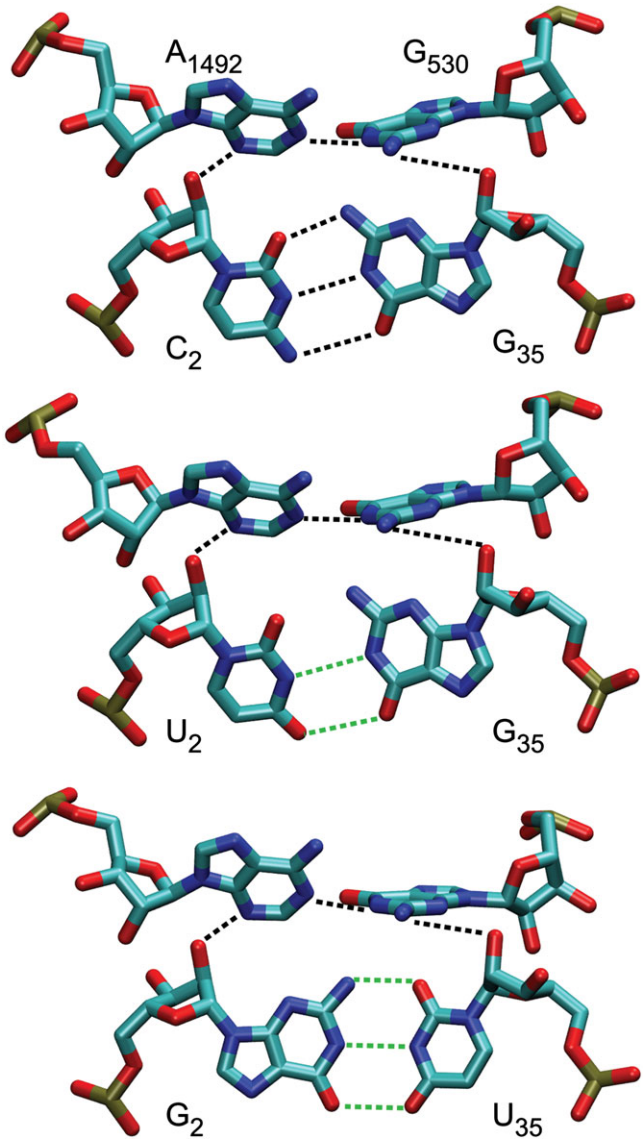

FIGURE 2 G_2•U_{35} and U_2•G_{35} mismatches mimic canonical G_2•C_{35} nucleotide pair. Predicted pairing of two mismatches and a canonical pair are shown. The black dotted lines indicate hydrogen bonds predicted by the authors of the papers describing each pair (Demeshkina et al., 2012; Rozov et al., 2018). Green dotted lines indicate hydrogen bonding that would require tautomeric shifts in one of the bases to accomplish. The images were created using the published data in the Protein Databank: 4V8E (G_2U_{35}), 6GSK (U_2•G_{35}), and 6GSJ (C_3•G_{35}). Images were created using VMD (Humphrey, Dalke, & Schulten, 1996) [Colour figure can be viewed at wileyonlinelibrary.com]

G•U, U•G, and U•U engaged the three rRNA nucleotides, the A•A and A•C pairs did not, indicating that the requirement for WC geometry is extremely strict. This, in turn, suggests that the restriction of misreading to G•U, U•G, and U•U (and presumably C•U) mismatches reflects their ability to engage with the A site and induce domain closure. The only mismatches that deviate from this model were identified by mass spectrometry in cases of nonsense suppression (Blanchet et al., 2014). These were all wobble position mismatches: A_3•G_{36}, G_3•G_{36}, and C_3•A_{36}. Structures with these mismatches are not available, but they are predicted to adopt non-WC geometry. The fact that errors requiring them are relatively efficient suggests that discrimination against alternate geometries is not as strict at the less monitored wobble position.

Whereas errors involving G•U mismatches are found in all cases tested, only one instance of a U•G wobble error has been reported (Z. Zhang et al., 2013), and the U•U and U•C mismatches are either mainly (or entirely, for U•C) restricted to the wobble position. The restriction of U•U and U•C errors may reflect the fact that the spacing between the two bases does not allow hydrogen bonds to form (Rozov et al., 2016). The lack of U•G errors must reflect a subtler structural effect or increased competition by the competing cognate tRNAs in the case of unobserved errors of this type.

4 | FUNDAMENTAL DIFFERENCES BETWEEN ERROR CORRECTION IN BACTERIA AND YEAST

Of particular relevance to this review is the fact that error correction in the *S. cerevisiae*, and perhaps all eukaryotes, is fundamentally unlike error correction in *E. coli* and perhaps all bacteria. There are several ways in which the bacterial and eukaryotic systems appear to differ.

First, errors are significantly more frequent overall in *E. coli* than in *S. cerevisiae*. Direct comparisons of error frequencies for $tRNA^{Lys}_{UUU}$ (Kramer et al., 2010; Kramer & Farabaugh, 2007) and $tRNA^{Glu}_{UUC}$ (Joshi et al., 2018; Manickam, Joshi, Bhatt, & Farabaugh, 2016) show that errors in *E. coli* average about threefold to fivefold greater than in *S. cerevisiae*. Because the frequency of errors is sensitive to the abundance of the cognate tRNA that competes for the misread codon, this difference could reflect a difference in the abundances of these competing cognate tRNAs or, for misreading of nonsense codons, differences in the availability of peptide release factors. There are approximately 7.5 tRNAs per *E. coli* ribosome (reviewed by Mackie, 2013) and about 25 per *S. cerevisiae* ribosome (reviewed by Warner, 1999). This difference could explain the greater accuracy in yeast because the greater stoichiometry of tRNAs to ribosomes might result in more frequent recruitment of cognate tRNAs. However, the frequency of errors is more likely to reflect relative differences in concentration between a particular cognate and near-cognate tRNA. A comparison based on available data for tRNA abundance (Dong, Nilsson, & Kurland, 1996; Percudani, Pavesi, & Ottonello, 1997) shows that the ratio of near-cognate and competing cognate tRNAs involved in middle position errors at AGG or GGA is greater in *S. cerevisiae* than in *E. coli* (Table 1). The greater ratios in *S. cerevisiae* would predict more effective competition by the near-cognate and therefore more frequent errors contrary to observation, the opposite of observed differences.

The second overall difference is that wobble misreading errors are much less frequent in *S. cerevisiae* than in *E. coli*, especially in the presence of error-inducing aminoglycoside antibiotics (Kramer & Farabaugh, 2007; Manickam et al., 2016). The difference between *E. coli* and *S. cerevisiae* is about 10-fold. Treatment with the antibiotic paromomycin increases wobble errors eightfold in *E. coli* but less than twofold in *S. cerevisiae*, increasing the difference to 40-fold. There is some support for tRNA competition causing this difference. In both species, the dosage of paromomycin was adjusted to be about twofold less than the amount leading to substantial decrease in growth, so we assume that the amount of paromomycin intracellular in each case is comparable. However, it is possible that at least some of the difference may reflect a difference in internal paromomycin concentration. The ratios of the relevant near-cognate and cognate tRNAs in *S. cerevisiae* are smaller than in *E. coli* (Table 1). This difference is consistent with fewer wobble position errors in *S. cerevisiae* as observed. The differences in ratios are, however, only about twofold in each case, and it is unclear if that is sufficient to explain the large observed differences in misreading frequency.

Another possible explanation of these differences is tRNA modification. Both $tRNA^{Lys}_{UUU}$ and $tRNA^{Glu}_{UUC}$ are modified at U_{34}, the wobble position of the anticodon, but in *E. coli*, the modification is to 5-methylaminomethyl-2-thiouridine (mnm^5s^2U), whereas in *S. cerevisiae*, it is 5-methoxylcarbonylmethyl-2-thiouridine (mcm^5s^2U). Wobble xm^5 modifications were thought to restrict wobble errors generally (Yokoyama et al., 1985), which suggested a hypothesis where mcm^5 might restrict more efficiently than mnm^5 explaining the lower wobble errors in *S. cerevisiae*. However, their effect on

TABLE 1 Near-cognate to cognate tRNA ratios in *Escherichia coli* and *Saccharomyces cerevisiae*

Codon misread	Near-cognate tRNA	Cognate tRNA	Ratio (near-cognate/cognate)	
			Escherichia coli	*Saccharomyces cerevisiae*
AGG	$tRNA^{Lys}_{UUU}$	$tRNA^{Arg}_{CCU}$	4.6	7.0
AAY		$tRNA^{Asn}$	1.7	0.7
GGA	$tRNA^{Glu}_{UUC}$	$tRNA^{Gly}_{UCC}$	2.3	4.7
GGY		$tRNA^{Asp}$	2.0	0.9

misreading is more complicated. In bacteria, the presence of mnm^5 modification strongly decreases wobble errors by tRNA$^{Glu}_{UUC}$ but equally strongly increases errors by tRNA$^{Lys}_{UUU}$ (Manickam et al., 2016), whereas the presence of mcm^5 in *S. cerevisiae* tends to change accuracy by a smaller ratio and generally increases misreading (Joshi et al., 2018). Importantly, with equivalently modified tRNAs carrying the same s^2U_{34} wobble modification, the error frequency in *S. cerevisiae* is still much less than in *E. coli*, suggesting that the difference is not a result of wobble modification. It remains to be seen if modification at some other position might explain the difference, but there are few other differences in modification of the *E. coli* and *S. cerevisiae* tRNAs and none with an obvious role in decoding.

The remaining explanation is that some fundamental difference between the *E. coli* and *S. cerevisiae* translation system causes the difference in error frequency. This could be a change in an rRNA or ribosomal protein, either primary sequence or modification, or a difference between EF-Tu and its cognate eEF-1A. It is attractive to suppose that eukaryotes generally and humans in particular show a similar suppression of wobble misreading, making *S. cerevisiae* a better model than *E. coli* for accuracy in humans. The availability of highly quantitative and reproducible reporter systems will be an important tool in the genetic dissection of the source of suppression of wobble misreading in *S. cerevisiae*.

5 | ARE THERE INVISIBLE NEAR-COGNATE "ERRORS" ELSEWHERE IN THE GENETIC CODE?

Analysis of the effect of modifications on misreading shows multiple examples where they increase misreading by stabilizing intrinsically unstable codon–anticodon complexes. Modifications to position 37, immediately adjacent to the first codon–anticodon base pair, stack on the codon–anticodon complex to stabilize it. tRNAs that recognize codons beginning in U (the top row of the conventional genetic code) and those recognizing codons beginning in A (the third row) are hypermodified at 37 (Figure 3). Most first row yeast tRNAs carry N^6-isopentenyladenosine (i6A; Nishimura, 1972), and third row tRNAs carry N^6-threonylcarbamoyladenosine (t6A; El Yacoubi et al., 2009; Srinivasan et al., 2011); bacterial tRNAs carry similar or identical modifications. Hyper modification of 37 is thought to stabilize the weak A•U or U•A pair to allow the tRNA to interact with the codon in the A site (Stuart, Koshlap, Guenther, & Agris, 2003), and the effect of these modifications is to increase misreading (Joshi et al., 2018; Lamichhane et al., 2013; Manickam et al., 2016). Other tRNAs with G•C or C•G base pairs at the first codon position have much smaller modifications of nucleotide 37 (m^1G or m^2A) or none at all, implying that the stability of the codon•anticodon complex is sufficient without the stabilizing effect of nucleotide 37 modifications. This implies that misreading by these tRNAs would be naturally high. Most tRNAs cannot form the G•U pairs that mimic WC pairs and those that cannot do not generate significant misreading (Z. Zhang et al., 2013). However, in every four-member codon box, there is the possibility for misreading by U_3•U_{34} or C_3•U_{34} mismatches. Is it possible that significant near-cognate decoding might occur by this mechanism across the genetic code?

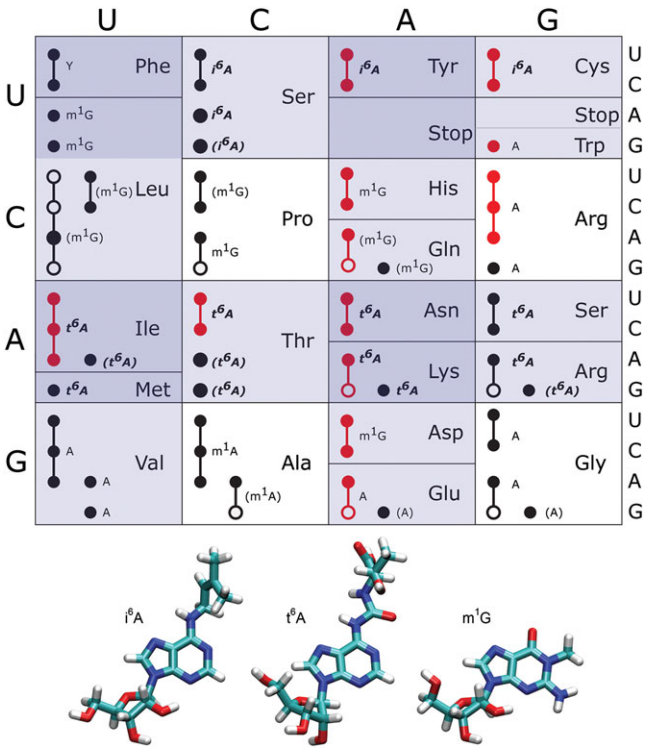

FIGURE 3 Split codon boxes are limited to weak codons with A or U at the first two codon positions. The *Saccharomyces cerevisiae* cytoplasmic tRNA complement is shown as in Figure 1 with the identity of nucleotide 37 in each tRNA shown. Columns and rows of the standard genetic code are highlighted in blue to indicate the presence of A or U in the first or second codon positions; where both positions are A/U the highlight is doubled. Note the coincidence of presence of A/U in these positions, the frequent use of hypermodification and the presence of split codon boxes involving more than one amino acid specified in a four-codon box. The structure of the three main nucleotide 37 modifications—hypermodified i^6A and t^6A and non-hypermodified m^1G—are shown below the codon table (generated using VMD from data files found at the Modomics web site; Boccaletto et al., 2018) [Colour figure can be viewed at wileyonlinelibrary.com]

In bacteria, the wobble U of tRNAs decoding fourfold degenerate codon boxes are commonly modified to uridine 5-oxyacetic acid (cmo^{5}U). The function of this modification is to allow the tRNA to recognize as many as all four codons (Nasvall, Chen, & Bjork, 2004). The modification promotes broadened misreading by increasing tautomer formation to allow pairing with all four nucleotides (Weixlbaumer et al., 2007) and has been referred to as "almost-correct" (Kothe & Rodnina, 2007), reminiscent of the notion of WC mimicry proposed based on the recent X-ray crystallographic results. The predicted stability of mismatched codon–anticodon pairs across much of the genetic code suggests that near-cognate decoding might be quite common even in the absence of this modification in yeast. At the same time, most of the "stronger" tRNAs decode in four-member synonymous boxes so that near-cognate decoding of that sort would not cause a translational error. These "errors" being as common as we suspect could have constrained the evolution of the genetic code to limit split codon boxes to those recognized by tRNAs making the weakest codon–anticodon pair: the UUN box (Phe/Leu), the AUN box (Ile/Met), and boxes with a middle position A including UAN (Tyr/Stop),

CAN (His/Gln), AAN (Asn/Lys), and GAN (Asp/Glu). Ironically, that would suggest that these boxes where near-cognate decoding generates observable translational errors may in fact be the least prone to near-cognate decoding because of the instability of the near-cognate codon–anticodon complex.

6 | TO WHAT EXTENT IS THE PROTEOME REALLY STATISTICAL?

Early work on translational errors debunked the idea put forward by Pauling (1957) that the proteome might be significantly statistical with errors on the percent level across all proteins. However, the proteome was thought to be statistical at the level of several in 10,000 amino acids incorporated across the proteome. This effect would not be minor, as pointed out by Parker (1989), because even at the seemingly low average misreading frequency of 5×10^{-4}, a large minority of proteins would include at least one misincorporated amino acid. If the average were indeed correct and extended across the entire proteome, then this conclusion would be inevitable. However, a consideration of the distribution of errors shows that the average is misleading because most errors occur much less frequently (Joshi et al., 2018; Manickam et al., 2014; Z. Zhang et al., 2013; J. Zhang et al., 2016). The relatively high frequency of errors at a subset of codons suggests that the proteome has some statistical features; that is, because of misacylation and misreading the frequency of miscoding of proteins is large enough that a substantial fraction of the proteome diverges in sequence and therefore, possibly, in structure. Given that misreading has increasingly been shown to have a positive effect on the cell (for a review, see Ribas de Pouplana, Santos, Zhu, Farabaugh, & Javid, 2014), the susceptibility to misreading might provide a range of variability in enzyme function that might be useful to the cell. At an extreme, it is attractive to speculate that, as with programmed frameshifting and termination readthrough (Farabaugh, Qian, & Stahl, 2000; Namy, Rousset, Napthine, & Brierley, 2004), there might be instances in particular proteins where the frequency of misreading errors increases in response to sequence context to create a programmed misreading site. Alternatively, there may be physiological effects that exacerbate certain errors to increase amino acid substitutions across the proteome. The advantage of programmed events of this type could be, for example, the production of enzymes with alternative active site amino acids that produce enzymes with significantly different catalytic activity including variant substrate specificity or kinetic properties. These alternative properties could be exploited by the cell or engineered by researchers. We do not know enough about physiological or context effects on misreading to identify such events. The most likely way to identify such systems would be from whole proteome scale mass spectrometry. No candidate has yet been reported.

ORCID

Philip J. Farabaugh https://orcid.org/0000-0002-5658-7141

REFERENCES

Blanchet, S., Cornu, D., Argentini, M., & Namy, O. (2014). New insights into the incorporation of natural suppressor tRNAs at stop codons in Saccharomyces cerevisiae. *Nucleic Acids Research*, *42*, 10061–10072. https://doi.org/10.1093/nar/gku663

Blanchet, S., Cornu, D., Hatin, I., Grosjean, H., Bertin, P., & Namy, O. (2018). Deciphering the reading of the genetic code by near-cognate tRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, 3018–3023. https://doi.org/10.1073/pnas.1715578115

Boccaletto, P., Machnicka, M. A., Purta, E., Piatkowski, P., Baginski, B., Wirecki, T. K., … Bujnicki, J. M. (2018). MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Research*, *46*, D303–D307. https://doi.org/10.1093/nar/gkx1030

Brinkmann, U., Mattes, R. E., & Buckel, P. (1989). High-level expression of recombinant genes in *Escherichia coli* is dependent on the availability of the dnaY gene product. *Gene*, *85*, 109–114. https://doi.org/10.1016/0378-1119(89)90470-8

Bullwinkle, T., Lazazzera, B., & Ibba, M. (2014). Quality control and infiltration of translation by amino acids outside of the genetic code. *Annual Review of Genetics*, *48*, 149–166. https://doi.org/10.1146/annurev-genet-120213-092101

Calderone, T., Stevens, R., & Oas, T. (1996). High–level misincorporation of lysine for arginine at AGA codons in a fusion protein expressed in *Escherichia coli*. *Journal of Molecular Biology*, *262*, 407–412. https://doi.org/10.1006/jmbi.1996.0524

Crick, F., Barnett, L., Brenner, S., & Watts-Tobin, R. (1961). General nature of the genetic code for proteins. *Nature*, *192*, 1227–1232. https://doi.org/10.1038/1921227a0

Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., & Yusupova, G. (2012). A new understanding of the decoding principle on the ribosome. *Nature*, *484*, 256–259. https://doi.org/10.1038/nature10913

Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., & Yusupova, G. (2013). New structural insights into the decoding mechanism: Translation infidelity via a G· U pair with Watson–Crick geometry. *FEBS Letters*, *587*, 1848–1857. https://doi.org/10.1016/j.febslet.2013.05.009

Dong, H., Nilsson, L., & Kurland, C. G. (1996). Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *Journal of Molecular Biology*, *260*, 649–663. https://doi.org/10.1006/jmbi.1996.0428

El Yacoubi, B., Lyons, B., Cruz, Y., Reddy, R., Nordin, B., Agnelli, F., … de Crecy-Lagard, V. (2009). The universal YrdC/Sua5 family is required for the formation of threonylcarbamoyladenosine in tRNA. *Nucleic Acids Research*, *37*, 2894–2909. https://doi.org/10.1093/nar/gkp152

Farabaugh, P., Qian, Q., & Stahl, G. (2000). Programmed translational frameshifting, hopping and readthrough of termination codons. In J. Hershey, M. Mathews, & N. Sonenberg (Eds.), *Translational control of gene expression* (pp. 741–761). Cold Spring Harbor, NY: Cold Spring Harbor Press.

Fischer, N., Neumann, P., Bock, L. V., Maracci, C., Wang, Z., Paleskava, A., … Stark, H. (2016). The pathway to GTPase activation of elongation factor SelB on the ribosome. *Nature*, *540*, 80–85. https://doi.org/10.1038/nature20560

Hopfield, J. J. (1974). Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceedings of the National Academy of Sciences of the United States of America*, *71*, 4135–4139. https://doi.org/10.1073/pnas.71.10.4135

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*, 33–38. https://doi.org/10.1016/0263-7855(96)00018-5

Joshi, K., Trivedi, M. B., & Farabaugh, P. J. (2018). Codon-specific effects of tRNA anticodon loop modifications on translational misreading errors in the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res: City*, *46*, 10331–10339. https://doi.org/10.1093/nar/gky664

Kothe, U., & Rodnina, M. V. (2007). Codon reading by tRNAAla with modified uridine in the wobble position. *Molecular Cell*, *25*, 167–174. https://doi.org/10.1016/j.molcel.2006.11.014

Kramer, E., & Farabaugh, P. (2007). The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. *RNA*, *13*, 87–96. https://doi.org/10.1261/rna.294907

Kramer, E. B., Vallabhaneni, H., Mayer, L. M., & Farabaugh, P. J. (2010). A comprehensive analysis of translational missense errors in the yeast *Saccharomyces cerevisiae*. *RNA*, *16*, 1797–1808. https://doi.org/10.1261/rna.2201210

Lamichhane, T. N., Blewett, N. H., Crawford, A. K., Cherkasova, V. A., Iben, J. R., Begley, T. J., … Maraia, R. J. (2013). Lack of tRNA modification isopentenyl-A37 alters mRNA decoding and causes metabolic deficiencies in fission yeast. *Molecular and Cellular Biology*, *33*, 2918–2929. https://doi.org/10.1128/MCB.00278-13

Limbach, P. A., Crain, P. F., & McCloskey, J. A. (1994). Summary: The modified nucleosides of RNA. *Nucleic Acids Research*, *22*, 2183–2196. https://doi.org/10.1093/nar/22.12.2183

Loveland, A. B., Demo, G., Grigorieff, N., & Korostelev, A. A. (2017). Ensemble cryo-EM elucidates the mechanism of translation fidelity. *Nature*, *546*, 113–117. https://doi.org/10.1038/nature22397

Mackie, G. A. (2013). RNase E: At the interface of bacterial RNA processing and decay. *Nature Reviews. Microbiology*, *11*, 45–57. https://doi.org/10.1038/nrmicro2930

Manickam, N., Joshi, K., Bhatt, M. J., & Farabaugh, P. J. (2016). Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Research*, *44*, 1871–1881. https://doi.org/10.1093/nar/gkv1506

Manickam, N., Nag, N., Abbasi, A., Patel, K., & Farabaugh, P. J. (2014). Studies of translational misreading in vivo show that the ribosome very efficiently discriminates against most potential errors. *RNA*, *20*, 9–15. https://doi.org/10.1261/rna.039792.113

Namy, O., Rousset, J. P., Napthine, S., & Brierley, I. (2004). Reprogrammed genetic decoding in cellular gene expression. *Molecular Cell*, *13*, 157–168. https://doi.org/10.1016/S1097-2765(04)00031-0

Nasvall, S. J., Chen, P., & Bjork, G. R. (2004). The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAPro (cmo5UGG) promotes reading of all four proline codons in vivo. *RNA*, *10*, 1662–1673. https://doi.org/10.1261/rna.7106404

Ninio, J. (1975). Kinetic amplification of enzyme discrimination. *Biochimie*, *57*, 587–595. https://doi.org/10.1016/S0300-9084(75)80139-8

Nishimura, S. (1972). Minor components in transfer RNA: Their characterization, location, and function. *Progress in Nucleic Acid Research and Molecular Biology*, *12*, 49–85. https://doi.org/10.1016/S0079-6603(08)60659-5

Ogle, J. M., Murphy, F. V., Tarry, M. J., & Ramakrishnan, V. (2002). Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell*, *111*, 721–732. https://doi.org/10.1016/S0092-8674(02)01086-3

Parker, J. (1989). Errors and alternatives in reading the universal genetic code. *Microbiological Reviews*, *53*, 273–298.

Pauling, L. (1957). The probability of errors in the process of synthesis of protein molecules. In A. Birkhauser (Ed.), *Festschrift Arthur Stoll* (pp. 597–602). Basel: Birkhauser Verlag.

Pavlov, M. Y., Liljas, A., & Ehrenberg, M. (2017). A recent intermezzo at the Ribosome Club. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*. 20160185, https://doi.org/10.1098/rstb.2016.0185

Percudani, R., Pavesi, A., & Ottonello, S. (1997). Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *Journal of Molecular Biology*, *268*, 322–330. https://doi.org/10.1006/jmbi.1997.0942

Ribas de Pouplana, L., Santos, M. A., Zhu, J. H., Farabaugh, P. J., & Javid, B. (2014). Protein mistranslation: Friend or foe? *Trends in Biochemical Sciences*, *39*, 355–362. https://doi.org/10.1016/j.tibs.2014.06.002

Rodnina, M. V., Fischer, N., Maracci, C., & Stark, H. (2017). Ribosome dynamics during decoding. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*. 20160182, https://doi.org/10.1098/rstb.2016.0182

Rodnina, M. V., & Wintermeyer, W. (2001). Fidelity of aminoacyl-tRNA selection on the ribosome: Kinetic and structural mechanisms. *Annual Review of Biochemistry*, *70*, 415–435. https://doi.org/10.1146/annurev.biochem.70.1.415

Rozov, A., Demeshkina, N., Khusainov, I., Westhof, E., Yusupov, M., & Yusupova, G. (2016). Novel base-pairing interactions at the tRNA wobble position crucial for accurate reading of the genetic code. *Nature Communications*, *7*, 10457. https://doi.org/10.1038/ncomms10457

Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M., & Yusupova, G. (2015). Structural insights into the translational infidelity mechanism. *Nature Communications*, *6*, 7251. https://doi.org/10.1038/ncomms8251

Rozov, A., Westhof, E., Yusupov, M., & Yusupova, G. (2016). The ribosome prohibits the G•U wobble geometry at the first position of the codon-anticodon helix. *Nucleic Acids Research*, *44*, 6434–6441. https://doi.org/10.1093/nar/gkw431

Rozov, A., Wolff, P., Grosjean, H., Yusupov, M., Yusupova, G., & Westhof, E. (2018). Tautomeric G•U pairs within the molecular ribosomal grip and fidelity of decoding in bacteria. *Nucleic Acids Research*, *46*, 7425–7435. https://doi.org/10.1093/nar/gky547

Seetharam, R., Heeren, R. A., Wong, E. Y., Braford, S. R., Klein, B. K., Aykent, S., … Siegel, N. R. (1988). Mistranslation in IGF-1 during over-expression of the protein in *Escherichia coli* using a synthetic gene containing low frequency codons. *Biochemical and Biophysical Research Communications*, *155*, 518–523. https://doi.org/10.1016/S0006-291X(88)81117-3

Srinivasan, M., Mehta, P., Yu, Y., Prugar, E., Koonin, E. V., Karzai, A. W., & Sternglanz, R. (2011). The highly conserved KEOPS/EKC complex is essential for a universal tRNA modification, t6A. *The EMBO Journal*, *30*, 873–881. https://doi.org/10.1038/emboj.2010.343

Stuart, J. W., Koshlap, K. M., Guenther, R., & Agris, P. F. (2003). Naturally-occurring modification restricts the anticodon domain conformational space of tRNA (Phe). *Journal of Molecular Biology*, *334*, 901–918. https://doi.org/10.1016/j.jmb.2003.09.058

Thompson, R. C. (1988). EFTu provides an internal kinetic standard for translational accuracy. *Trends in Biochemical Sciences*, *13*, 91–93. https://doi.org/10.1016/0968-0004(88)90047-3

Warner, J. R. (1999). The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, *24*, 437–440. https://doi.org/10.1016/S0968-0004(99)01460-7

Weixlbaumer, A., Murphy, F. V., Dziergowska, A., Malkiewicz, A., Vendeix, F. A., Agris, P. F., & Ramakrishnan, V. (2007). Mechanism for expanding the decoding capacity of transfer RNAs by modification of uridines. *Nature Structural & Molecular Biology*, *14*, 498–502. https://doi.org/10.1038/nsmb1242

Wen, D., Vecchi, M. M., Gu, S., Su, L., Dolnikova, J., Huang, Y. M., … Meier, W. (2009). Discovery and investigation of misincorporation of serine at asparagine positions in recombinant proteins expressed in Chinese hamster ovary cells. *The Journal of Biological Chemistry*, *284*, 32686–32694. https://doi.org/10.1074/jbc.M109.059360

Woese, C. R. (1965). On the evolution of the genetic code. *Proceedings of the National Academy of Sciences of the United States of America*, *54*, 1546–1552. https://doi.org/10.1073/pnas.54.6.1546

Wohlgemuth, I., Pohl, C., & Rodnina, M. V. (2010). Optimization of speed and accuracy of decoding in translation. *The EMBO Journal*, *29*, 3701–3709. https://doi.org/10.1038/emboj.2010.229

Yokoyama, S., Watanabe, T., Murao, K., Ishikura, H., Yamaizumi, Z., Nishimura, S., & Miyazawa, T. (1985). Molecular mechanism of codon recognition by tRNA species with modified uridine in the first position of the codon. *Proceedings of the National Academy of Sciences of the United States of America*, *82*, 4905–4909. https://doi.org/10.1073/pnas.82.15.4905

Yu, X. C., Borisov, O. V., Alvarez, M., Michels, D. A., Wang, Y. J., & Ling, V. (2009). Identification of codon-specific serine to asparagine mistranslation in recombinant monoclonal antibodies by high-resolution mass spectrometry. *Analytical Chemistry*, *81*, 9282–9290. https://doi.org/10.1021/ac901541h

Zhang, J., Ieong, K. W., Mellenius, H., & Ehrenberg, M. (2016). Proofreading neutralizes potential error hotspots in genetic code translation by transfer RNAs. *RNA*, *22*, 896–904. https://doi.org/10.1261/rna.055632.115

Zhang, Z., Shah, B., & Bondarenko, P. V. (2013). G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. *Biochemistry*, *52*, 8165–8176. https://doi.org/10.1021/bi401002c

How to cite this article: Joshi K, Cao L, Farabaugh PJ. The problem of genetic code misreading during protein synthesis. *Yeast*. 2019;36:35–42. https://doi.org/10.1002/yea.3374